

Original-Titel:

Comparing the number and relevance of false activations between 2 artificial intelligence computer-aided detection systems: the NOISE study

Autoren:

Marco Spadaccini, Cesare Hassan, Ludovico Alfarone, Leonardo Da Rio, Roberta Maselli, Silvia Carrara, Piera Alessia Galtieri, Gaia Pellegatta, Alessandro Fugazza, Glenn Koleth, James Emmanuel, Andrea Anderloni, Yuichi Mori, Michael B. Wallace, Prateek Sharma, Alessandro Repici, Gastrointest Endosc, Gastrointest Endosc 2022; 95:975-81,(Clinical trial registration number: NCT04399590.)

Kommentar:

Prof. Dr. Jörg G. Albert, Stuttgart, 19.08.2022

Die Computer-unterstützten Detektionsverfahren (CADE), die auf sogenannten ‚künstlichen Intelligenzverfahren‘ (AI) beruhen, können mit einer hohen Sensitivität Kolonpolypen entdecken und haben in randomisierten Studien eine erhöhte Polypendetektionsrate und verringerte ‚miss-rate‘ erzielen können. Inzwischen sind mehrere CADe-Systeme zugelassen und am Markt erhältlich. Im klinischen Einsatz kann die falsch-positive (FP) Aktivierung allerdings eine wesentliche Einschränkung des erfolgreichen Einsatzes darstellen und die Routine einer Untersuchung stören.

Ziel der vorliegenden Studie war es daher, zwei zur Verfügung stehende CADe-Systeme anhand einer systematischen Erfassung der FP-Befunde zu vergleichen. Hierbei wurden die Anzahl der CADe-FP Aktivierung, die Ursache und die zusätzliche Zeit, die diese FP den Untersucher gekostet haben, untersucht. Es wurden zwei Videosammlungen à 40 Videos jeweils mit dem einen und dem anderen Verfahren geprüft. (CADE A und CADe B). Es wurde entsprechend eine Klassifizierung, die bereits in einer vorherigen Studie eingesetzt worden war, herangezogen (die ‚NOISE‘ Klassifikation): u.a. war hier die klinische Relevanz nach der Dauer der Beschäftigung des Untersuchers mit dem Artefakt, das zur FP-Aktivierung geführt hat, eingeordnet (keine Relevanz, 0 Sekunden; milde Relevanz, 1s; moderate, 1 – 3s; schwere >3s).

Im Ergebnis zeigte sich eine FP-Rate von $25,5 \pm 12,2$ FPs pro Koloskopie bei CADe A (1021 FP Aktivierungen) gegenüber einer FP-Rate von $25,7 \pm 13,2$ FPs pro Koloskopie (1028 FP Aktivierungen) bei CADe B; $p = 0,53$. Als Ursache wurden in vergleichbarer Häufigkeit missdeutete Strukturen der Darmwand sowie Darminhalt/Stuhl gefunden. Auch die zusätzlich aufgewendete Zeit des Untersuchers war bei beiden Systemen vergleichbar und betrug $0,2 \pm 0,9$ Sekunden vs. $0,2 \pm 0,8$ Sekunden pro falscher Aktivierung, wobei nur ein Teil der FP Befunde. Es erforderten allerdings nur 65 FP der 1021 bei CADe A und 72 der 1028 FP bei CADe B eine zusätzliche Inspektion, bei der jeweils ca. $\frac{1}{4}$ eine milde und eine schwere Relevanz, und jeweils $\frac{1}{2}$ eine moderate Relevanz nach ‚NOISE‘ aufwiesen.

Die Wissenschaftler schlussfolgerten, dass die beiden CADe-Systeme vergleichbare Einschränkungen aufwiesen und dass ‚NOISE‘ ein Benchmark für zukünftige CADe Entwicklungen darstellen könnte. Zwar wurde eine relativ hohe Anzahl an FP Aktivierungen beobachtet, diese gingen aber mit einer klinisch vernachlässigbaren Einschränkung einher, zumindest was die Untersuchungszeit betrifft. Damit ergab sich keine relevante Einschränkung der CADe-Systeme für den klinischen Einsatz.

Literatur

Spadaccini M, Iannone A, Maselli R, et al. Computer-aided detection versus advanced imaging for detection of colorectal neoplasia: a systematic review and network meta-analysis. *Lancet Gastroenterol Hepatol* 2021;6:793-802.